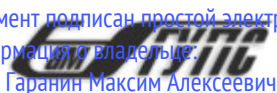


Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Гаранин Максим Алексеевич
Должность: Ректор
Дата подписания: 03.10.2023 17:04:00
Уникальный программный ключ:
7708e3a47e66a8ee02711b298d7c78bd1e40bf88



МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ АГЕНТСТВО ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
САМАРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПУТЕЙ СООБЩЕНИЯ

23.05.03 Подвижной состав железных дорог

Приложение
к рабочей программе дисциплины

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

ПРАКТИКУМ ПО МАШИННОМУ ОБУЧЕНИЮ

(наименование дисциплины(модуля))

Направление подготовки / специальность

23.05.01 Наземные транспортно-технологические средства

(код и наименование)

Направленность (профиль)/специализация

Подъемно-транспортные, строительные, дорожные средства и оборудование

(наименование)

Содержание

1. Пояснительная записка.
2. Типовые контрольные задания или иные материалы для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих уровень сформированности компетенций.
3. Методические материалы, определяющие процедуру и критерии оценивания сформированности компетенций при проведении промежуточной аттестации.

1. Пояснительная записка

Цель промежуточной аттестации – оценивание промежуточных и окончательных результатов обучения по дисциплине, обеспечивающих достижение планируемых результатов освоения образовательной программы.

Формы промежуточной аттестации: Зачёт, 8 семестр

Перечень компетенций, формируемых в процессе освоения дисциплины

Код и наименование компетенции	Код индикатора достижения компетенции
ОПК-2: Способен решать профессиональные задачи с использованием методов, способов и средств получения, хранения и переработки информации; использовать информационные и цифровые технологии в профессиональной деятельности	ОПК-2.4: Решает задачи в области профессиональной деятельности, используя перспективные методы машинного обучения

Результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Оценочные материалы
ОПК-2.4: Решает задачи в области профессиональной деятельности, используя перспективные методы машинного обучения	Обучающийся знает: - основные классы научно-технических задач в области своей профессиональной деятельности, решаемые методами машинного обучения; - основные классы интеллектуальных информационных систем, необходимые в процессе решения научно-технических задач в области своей профессиональной деятельности;	Вопросы (№1 - №30)
	Обучающийся умеет: - использовать методы искусственного интеллекта для решения прикладных задач;	Задания (№1 - №30)
	Обучающийся владеет: - навыками решения научно-технических задач в области своей профессиональной деятельности с использованием технологий анализа данных и машинного обучения.	Задания (№ 36- №40)

Промежуточная аттестация (экзамен) проводится в одной из следующих форм:

- 1) ответ на билет, состоящий из теоретических вопросов и практических заданий;
- 2) выполнение заданий в ЭИОС СамГУПС.

Промежуточная аттестация (зачет) проводится в одной из следующих форм:

- 1) собеседование;
- 2) выполнение заданий в ЭИОС СамГУПС.

2. Типовые¹ контрольные задания или иные материалы для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих уровень сформированности компетенций

2.1. Типовые вопросы (тестовые задания) для оценки знаниевого образовательного результата

Проверяемый образовательный результат:

Задание 1 Машинное обучение —	
а) специализированное программное решение (или набор решений), которое включает в себя все инструменты для извлечения закономерностей из сырых данных	
б) эта группировка объектов (Наблюдений, событий) на основанных, описывающих свойства объектов	
в) набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной набор данных, и соответствующий ему правильный выходной результат	
г) подразделение искусственного интеллекта изучающий методы построения алгоритмов, способных обучаться на данных	
Задание 2 Среди предложенных задач машинного обучения укажите задачи регрессии	
а) Поиск негативных отзывов на фильм на сайте кинокомпании	
б) Алгоритм фильтрации спама	
в) Предсказание срока окупаемости проекта	
г) Предсказание рыночной стоимости квартиры	
д) Поиск мошеннических транзакций	
Задание 3 К библиотекам анализа данных в Python относятся ...	
а) Tkinter	
б) Pandas	
в) NumPy	
г) Matplotlib	
Задание 4 Какие из перечисленных команд вернут данные для России и Зимбабве?	
а) <code>res.query('country == ["Russia","Zimbabwe"]')</code>	
б) <code>res.query('country == "Russia" "Zimbabwe")</code>	
в) <code>res.query('country == "Russia" country == "Zimbabwe")</code>	
г) <code>res.query('country == "Russia" & country == "Zimbabwe")</code>	
д) <code>res[res.country == "Russia" res.country == "Zimbabwe"]</code>	
е) <code>res[(res.country == "Russia") (res.country == "Zimbabwe")]</code>	
Задание 5 Отметьте верные утверждения об алгоритме случайного леса - Random Forest	
а) Итоговым предсказанием модели является предсказание случайного дерева	
б) Параметры для каждого дерева (глубина, минимальное число образцов в листе и т.д.) выбираются случайно	
в) Каждое дерево в лесу получает случайный поднабор данных	
г) Число деревьев в лесу выбирается случайным образом	
д) Предсказание модели - усреднённые предсказания деревьев	
Задание 6 Какой атрибут отвечает за минимальное число образцов в листе?	
а) <code>min_samples_leaf</code>	
б) <code>max_depth</code>	
в) <code>min_samples_split</code>	
г) <code>min_impurity_decrease</code>	
Задание 7 Какой атрибут отвечает за минимальное число образцов в узле для разделения?	
а) <code>min_samples_leaf</code>	
б) <code>max_depth</code>	

¹ Приводятся типовые вопросы и задания. Оценочные средства, предназначенные для проведения аттестационного мероприятия, хранятся на кафедре в достаточном для проведения оценочных процедур количестве вариантов. Оценочные средства подлежат актуализации с учетом развития науки, образования, культуры, экономики, техники, технологий и социальной сферы. Ответственность за нераспространение содержания оценочных средств среди обучающихся университета несут заведующий кафедрой и преподаватель – разработчик оценочных средств.

2.2. Типовые задания для оценки навыкового образовательного результата

Проверяемый образовательный результат:

Код и наименование индикатора достижения компетенции	Образовательный результат
ОПК-2.4: Решает задачи в области профессиональной деятельности, используя перспективные методы машинного обучения	Обучающийся умеет: - использовать методы искусственного интеллекта для решения прикладных задач;

Приложение 2

Задание №1

Для ответа на этот вопрос вам понадобится набор данных imdb из библиотеки keras
`from keras.datasets import imdb`

Определите, какое слово стоит на 1000 месте (имеет индекс 1000) в частотном словаре датасета IMDb?

Запишите ответ:

1) Ответ:

Задание №2

Для ответа на этот вопрос вам понадобится датасет imdb из библиотеки keras
`from keras.datasets import imdb`

Определите, какое слово стоит на 3355 месте (имеет индекс 3355) в частотном словаре датасета IMDb?

Запишите ответ:

1) Ответ:

Задание №3

Для ответа на этот вопрос вам понадобится датасет imdb из библиотеки keras
`from keras.datasets import imdb`

Определите на каком месте в списке самых часто встречающихся слов в наборе данных IMDb стоит слово "joker" (впишите полученный индекс)?

Запишите ответ:

1) Ответ:

Задание №4

Для ответа на этот вопрос вам понадобится датасет imdb из библиотеки keras

```
from keras.datasets import imdb
```

Определите на каком месте в списке самых часто встречающихся слов в наборе данных IMDb стоит слово "study" (впишите полученный индекс)?

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №5

В какой стране значение соотношения между количеством вакцинированных и общей численностью населения на текущую дату в стране максимально? Поле Daily vaccinations per million.

Источник данных - country_vaccinations.csv

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №6

В какой строке (укажите индекс) находится максимальное значение по полю daily_vaccinations_per_million для стран Италия, Испания и Португалия.

Ответ - одно число, индекс строки.

Источник данных - country_vaccinations.csv

Запишите число:

1)	Ответ:	
----	--------	--

Задание №7

В какой строке указаны максимальные показатели по daily_vaccinations_per_million для

Аргентины, России и Уэльса.

Укажите номера строк через пробел. Например: 45 678 2223

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №8

Укажите название библиотеки позволяющей строить диаграммы рассеивания, графики и гистограммы в Python

Запишите ответ:

1)

Ответ:

Задание №9

Имеется набор данных о работе датчиков в грузовых вагонах поезда. Необходимо по датчикам автоматически разделять вагоны на полностью заполненные и пустые.

Напишите команду, создающую дерево решений dt для этого набора.

Запишите ответ:

1)

Ответ:

Задание №10

Какое значение имеет параметр daily_vaccinations для строки с индексом 3567?

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №11

Какое значение имеет параметр daily_vaccinations строки датасета country_vaccinations с индексом 3367?

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №12

Какое значение имеет параметр daily_vaccinations строки с индексом 25?

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №13

Какое значение имеет параметр people_fully_vaccinated_per_hundred для строки датасета с индексом 4036?

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №14

Какое значение имеет параметр people_vaccinated для последней строки датасета country_vaccinations?

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №15

Какое значение имеет параметр total_vaccinations_per_hundred для строки датасета с

индексом 4036?

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №16

Работа с файлом country_vaccinations.csv

Какое число указано в строке с индексом "3000" в столбце daily_vaccinations_per_million?

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №17

При обучении дерева решений на обучающей выборке, получена 100% точность классификатора, однако на тестовых классификатор показал точность около 50%. Что может быть причиной такого результата?

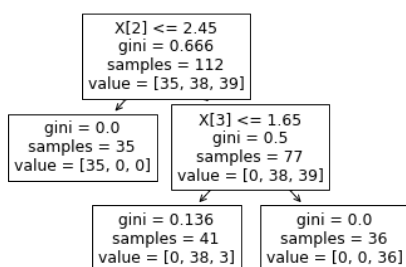
Запишите ответ:

1)

Ответ:

Задание №18

Перед вами дерево решений. Укажите самый высокий коэффициент загрязнения в листе (не во внутреннем узле)



Запишите ответ:

1)

Ответ:

Задание №19

Отберите данные за 5 февраля 2021 года. В какой стране количество привитых за день было максимально?

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №20

Отсортируйте данные о вакцинации в Испании по полю daily_vaccinations. Используйте сортировку по убыванию. Укажите значение поля daily_vaccinations в третьей строке.

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №21

Посчитайте в скольких наблюдениях в качестве вакцины использовался "Sputnik V" и "Pfizer".

Укажите оба числа через пробел. Сначала меньшее, потом большее.

Чтобы найти страны, в которых применялась вакцина, можно воспользоваться методом `Series.str.contains`

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №22

Для ответа на этот вопрос вам понадобится датасет imdb из библиотеки keras

```
from keras.datasets import imdb
```

Определите какое из слов lucky или drama чаще встречается в датасете IMDb? УВ качестве ответа укажите абсолютную (по модулю) разницу между позициями этих слов в частотном словаре?

Запишите число:

1)

Ответ:

Задание №23

Сгруппируйте данные по дате наблюдения. Посчитайте сумму по полю daily_vaccinations. Укажите второе по величине значение.

Ответ - сумма, а не дата.

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №24

Сгруппируйте данные по странам, для каждой страны вычислите сумму по полю "daily_vaccinations", отсортируйте полученные данные по убыванию. Укажите значение в третьей строке.

Ответ - одно число.

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №25

Сгруппируйте данные по странам. Укажите страну, которая находится на втором месте

по максимальному числу полностью вакцинированных (people_fully_vaccinated).

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №26

Сколько людей было вакцинировано в России 10го февраля? Поле daily_vaccinations.

Источник данных - country_vaccinations.csv

Запишите ответ:

1)

Ответ:

Задание №27

Создайте случайные лес rf состоящий из 15 деревьев. В ответе укажите строчку кода, отвечающую за его создание.

Запишите число:

1)

Ответ:

Задание №28

Имеется матрица ошибок:

Прогнозируемый	1	2	3	All
Фактический				
1	189	894	3028	4111
2	912	7195	30044	38151
3	4022	36797	223878	264697
All	5123	44886	256950	306959

Посчитайте показатель recall для класса с меткой 3.

Запишите число:

1)

Ответ:

Задание №29

Имеется матрица ошибок:

Прогнозируемый	1	2	3	All
Фактический				
1	189	894	3028	4111
2	912	7195	30044	38151
3	4022	36797	223878	264697
All	5123	44886	256950	306959

Посчитайте показатель precision для класса с меткой 2.

Запишите число:

1)

Ответ:

Задание №30

Загрузите встроенный набор данных digits.

```
from sklearn.datasets import load_digits
```

```
digs = load_digits()
```

Разделите данные на тестовый и тренировочный наборы в пропорции 75% и 25%.

Создайте дерево решений.

Выявите лучшие параметры дерева решений на следующих данных: максимальная глубина дерева от 2 до 10, минимальное количество выборок, необходимых создания листа (внешнего узла) от 4 до 6. В ответ запишите полученные показатели через пробел, сначала максимальную глубину, затем минимальное количество выборок, необходимых для создания узла.

Запишите ответ:

1)

Ответ:

Ответы:

#1 (1 б.)

Ответ = 'secret'

#2 (1 б.)

Ответ = 'loser'

#3 (1 б.)

Ответ = 6196

#4 (1 б.)

Ответ = 2075

#5 (1 б.)

Ответ = Falkland Islands

#6 (1 б.)

Ответ = 2034

#7 (1 б.)	Ответ = 143 3154 4024
#8 (1 б.)	Ответ = Python
#9 (1 б.)	Ответ = dt = DecisionTreeClassifier()
#10 (1 б.)	Ответ = 70598.0
#11 (1 б.)	Ответ = 2383.0
#12 (1 б.)	Ответ = 24.0
#13 (1 б.)	Ответ = 1.58
#14 (1 б.)	Ответ = 4041.0
#15 (1 б.)	Ответ = 29.16
#16 (1 б.)	Ответ = 2165.0
#17 (1 б.)	Ответ = Переобучение
#18 (1 б.)	Ответ = 0,136
#19 (1 б.)	Ответ = China
#20 (1 б.)	Ответ = 74783.0
#21 (1 б.)	Ответ = 391 3126
#22 (1 б.)	Ответ = 1589
#23 (1 б.)	Ответ = 6085565.0
#24 (1 б.)	Ответ = 17868335.0
#25 (1 б.)	Ответ = Israel
#26 (1 б.)	Ответ = 103571.0
#27 (1 б.)	Ответ = rf = RandomForestClassifier(n_estimators = 15)
#28 (1 б.)	Ответ = 0.845790
#29 (1 б.)	Ответ = 0.160295
#30 (1 б.)	Ответ = 9 4

области профессиональной деятельности, перспективные методы машинного обучения	- навыками решения научно-технических задач в области своей профессиональной деятельности с использованием технологий анализа данных и машинного обучения.
<p>Задание 31</p> <p>1. Скачайте набор данных о поездах и транспортной сети Франции. https://www.kaggle.com/datasets/gatandubuc/public-transport-traffic-data-in-france. Прочтите данные из файла Regularities_by_liaisons_Trains_France.csv в датафрейме rb. Получите распределение числовых величин датафрейма rb используя базовые методы визуализации библиотеки pandas. Определите графически в каком месяце было больше всего отмененных поездов? Связаны ли пики данных с общим количеством запланированных поездок? Исследуйте графически корреляцию между числовыми признаками в наборе данных rb. Постройте тепловую карту.</p> <p>Задание 32</p> <p>7. Скачайте данные из набора https://www.kaggle.com/code/ambaniverma/uk-traffic-accidents/data Используя средства визуализации, проведите исследование, есть ли параметры, которые наиболее сильно влияют на дорожные происшествия? В наборе не наблюдается сильных положительных корреляций, с погодой, местностью, возрастом водителя и т.д. Определите количество инцидентов в зависимости от зоны ограничения скорости.</p> <p>Задание 33</p> <p>Зарегистрируйтесь в системе организации конкурсов по исследованию данных Kaggle. Скачайте набор данных о поездах и транспортной сети Франции. https://www.kaggle.com/datasets/gatandubuc/public-transport-traffic-data-in-france В разделе Details изучите информацию о файле Regularities_by_liaisons_Trains_France.csv. Ответьте на вопрос, в каком случае поезд считается прибывшим вовремя? Отсортируйте годы ведения статистики по общему количеству поездов, опоздавших более, чем на 15 минут. Выведите количество отмененных поездов по каждой станции отправления.</p>	

2.3. Перечень вопросов для подготовки обучающихся к промежуточной аттестации

1. Машинное обучение.
2. Глубокое обучение.
3. Обучение с учителем и без учителя.
4. Основные понятия анализа данных. Данные, модель, обработка данных.
5. Методологические принципы анализа данных.
6. Цели, этапы, методы и техники анализа данных.
7. Фильтрация, группировка и агрегация данных.
8. Визуализация данных.
9. Визуализация данных в Python.
10. Визуализация в Pandas.
11. Библиотеки Seaborn и Matplotlib
12. Решающие деревья: основные понятия.
13. Задача классификации.
14. Деревья решений. Основные параметры дерева.
15. Дерево решений в задачи регрессии.
16. Метод ближайших соседей.
17. Концепция случайного леса.
18. Случайная выборка тренировочных образцов.
19. Усреднение прогнозов.
20. Проблема переобучения.
21. Типы ошибок: true negative, false positive.
22. Метрики качества модели: Precision, Recall, F1 score.
23. Тестирование ROC AUC.
24. Нейронные сети.
25. Искусственные нейроны.
26. Функции активации нейрона.
27. Перцептрон.
28. Обучение нейрона.
29. Градиентный спуск.
30. Дробление шага при градиентном спуске.
31. Стохастический градиентный спуск.
32. Метод наискорейшего спуска.
33. Нейронная сеть.
34. Однослойная модель.
35. Сверточные нейронные сети.
36. Свёрточный слой, слой подвыборки, полносвязный слой.
37. Целевая функция.
38. Задачи распознавания образов.
39. Классификация объектов.
40. Поиск изображения по образцу.

3. Методические материалы, определяющие процедуру и критерии оценивания сформированности компетенций при проведении промежуточной аттестации

Критерии формирования оценок по ответам на вопросы, выполнению тестовых заданий

- оценка **«отлично»** выставляется обучающемуся, если количество правильных ответов на вопросы составляет 100 – 90% от общего объема заданных вопросов;
- оценка **«хорошо»** выставляется обучающемуся, если количество правильных ответов на вопросы – 89 – 76% от общего объема заданных вопросов;
- оценка **«удовлетворительно»** выставляется обучающемуся, если количество правильных ответов на тестовые вопросы – 75–60 % от общего объема заданных вопросов;
- оценка **«неудовлетворительно»** выставляется обучающемуся, если количество правильных ответов – менее 60% от общего объема заданных вопросов.

Критерии формирования оценок по результатам выполнения заданий

«Отлично» – ставится за работу, выполненную полностью без ошибок и недочетов.

«Хорошо» – ставится за работу, выполненную полностью, но при наличии в ней не более одной негрубой ошибки и одного недочета, не более трех недочетов.

«Удовлетворительно» – ставится за работу, если обучающийся правильно выполнил не менее 2/3 всей работы или допустил не более одной грубой ошибки и двух недочетов, не более одной грубой и одной негрубой ошибки, не более трех негрубых ошибок, одной негрубой ошибки и двух недочетов.

«Неудовлетворительно» – ставится за работу, если число ошибок и недочетов превысило норму для оценки «удовлетворительно» или правильно выполнено менее 2/3 всей работы.

Виды ошибок:

- *грубые ошибки: незнание основных понятий, правил, норм; незнание приемов решения задач; ошибки, показывающие неправильное понимание условия предложенного задания.*

- *негрубые ошибки: неточности формулировок, определений; нерациональный выбор хода решения.*

- *недочеты: нерациональные приемы выполнения задания; отдельные погрешности в формулировке выводов; небрежное выполнение задания.*

Критерии формирования оценок по зачету

«Зачтено» – студент демонстрирует знание основных разделов программы изучаемого курса: его базовых понятий и фундаментальных проблем; приобрел необходимые умения и навыки, освоил вопросы практического применения полученных знаний, не допустил фактических ошибок при ответе, достаточно последовательно и логично излагает теоретический материал, допуская лишь незначительные нарушения последовательности изложения и некоторые неточности.

«Незачтено» – выставляется в том случае, когда студент демонстрирует фрагментарные знания основных разделов программы изучаемого курса: его базовых понятий и фундаментальных проблем. У экзаменуемого слабо выражена способность к самостоятельному аналитическому мышлению, имеются затруднения в изложении материала, отсутствуют необходимые умения и навыки, допущены грубые ошибки и незнание терминологии, отказ отвечать на дополнительные вопросы, знание которых необходимо для получения положительной оценки.